

Lista 4

1. O Chateau Latour é amplamente reconhecido como uma das maiores vinícolas do mundo, com uma rica história que remonta pelo menos a 1638. O Grand Vin de Château Latour é um vinho de incrível poder e longevidade. Em uma degustação em Nova York em abril de 2000, as safras de Latour de 1863 e 1899 foram classificadas, juntamente com as safras de 1945 e de 1961, como as melhores em um conjunto de 39 safras que vão de 1863 a 1999 (Wine Spectator, 31 de agosto de 2000). A qualidade de uma determinada safra do Chateau Latour tem um enorme impacto no preço. Por exemplo, em março de 2007, a safra 1997 do Chateau Latour poderia ser comprada por US\$159 por garrafa, enquanto a safra 2000 do Chateau Latour custa pelo menos US\$700 por garrafa (www.wine-searcher.com).

Embora muitos estudos tenham identificado que o momento da colheita das uvas tem um efeito importante na qualidade do vintage (ano em que as uvas foram colhidas), e a qualidade melhora quanto mais cedo é a colheita. Uma questão de interesse menos explorada é o efeito da chuva indesejada no tempo do vintage sobre a qualidade do vinho ícone como Chateau Latour. Essa questão aborda este problema.

O site do Chateau Latour (www.chateau-latour.com) fornece uma rica fonte de dados. Em particular, dados sobre a qualidade de cada safra, datas de colheita e clima na época da colheita foram obtidos no local para as safras de 1961 a 2004. Um exemplo das informações sobre clima na época da colheita é dado abaixo para a safra de 1994:

A colheita começou no dia 13 de setembro e durou até dia 29, frequentemente interrompida por tempestades. Mas, surpreendentemente, o efeito de diluição nas uvas foi muito limitado....
(<http://www.chateau-latour.com/commentaires/1994uk.html>)

Cada safra foi classificada como tem tido "chuva indesejada na colheita" (por exemplo, a safra de 1994) ou não (por exemplo, a safra de 1996) com base nas informações embaixo. Assim, os dados consistem em:

Vintage = ano que as uvas foram colhidas

Qualidade (Quality)- em uma escala de 1 (pior) a 5 (melhor) com alguns meio pontos

Fim da colheita (End of Harvest) - medido como o número de dias desde 31 de agosto

Chuva (rain)- uma variável dummy para chuva indesejada na colheita = 1 se sim.

Os dados podem ser encontrados no arquivo latour.csv. O **primeiro modelo** considerado foi:

$$\text{Quality} = \beta_0 + \beta_1 \text{End of Harvest} + \beta_2 \text{Rain} + \beta_3 \text{End of Harvest} \times \text{Rain} + e$$

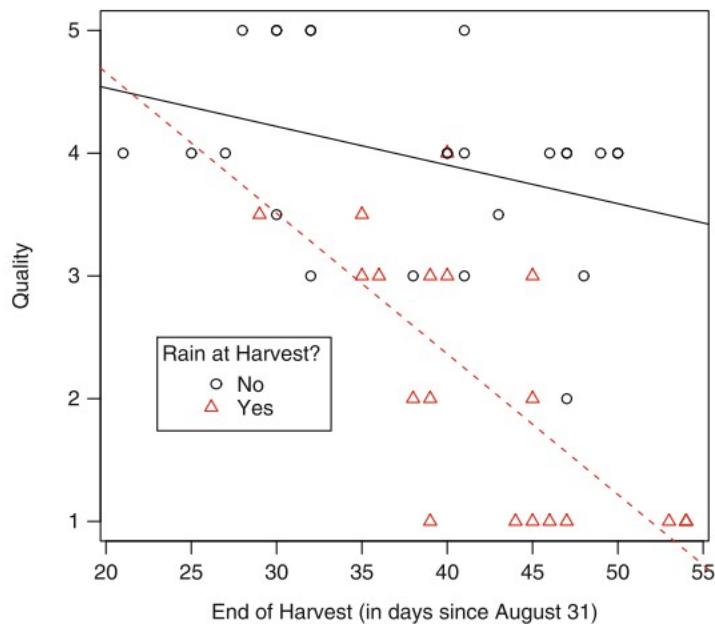
Um gráfico dos dados e as duas linhas de regressão do modelo podem ser encontradas embaixo e o Código R.

(a) Mostre que o coeficiente do termo de interação no modelo é estatisticamente significativo. Em outras palavras, mostrar que a taxa de mudança na qualidade depende se houve alguma chuva indesejada no vintage.

(b) Faça análise de variância

(c) Realizar diagnóstico do modelo

```
#Output da Regressao
mfull <- lm(Quality ~ EndofHarvest + Rain + Rain:EndofHarvest)
summary(mfull)
```



```
#Codigo Figura
y = Rain
par(mfrow=c(1,1))
plot(EndofHarvest,Quality,pch=y+1,col=y+1,xlab="End of Harvest (in days since August 31)")
abline(lsfrit(EndofHarvest[y==0],Quality[y==0]),lty=1,col=1)
abline(lsfrit(EndofHarvest[y==1],Quality[y==1]),lty=2,col=2)
legend(23, 2.5,legend=c("No","Yes"),pch=1:2,col=1:2,title="Rain at Harvest?")
```

2. O conjunto de dados reais nesta questão apareceu pela primeira vez em Half (1952). Os dados são fornecidos no arquivo Haldcement.txt e na Tabela embaixo. O centro de interesse é usar todas as técnicas de seleção de variáveis (estudas em aula), para escolher um subconjunto dos preditores para modelar Y. Em toda esta questão, vamos considerar que o primeiro modelo é para os dados

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e$$

- Discuta as técnicas AIC e BIC.
- Realizar diagnostico

Referência

Hald A (1952). Statistical theory with engineering applications. Wiley, New York.

Y	x_1	x_2	x_3	x_4
78.5	7	26	6	60
74.3	1	29	15	52
104.3	11	56	8	20
87.6	11	31	8	47
95.9	7	52	6	33
109.2	11	55	9	22
102.7	3	71	17	6
72.5	1	31	22	44
93.1	2	54	18	22
115.9	21	47	4	26
83.8	1	40	23	34
113.3	11	66	9	12
109.4	10	68	8	12